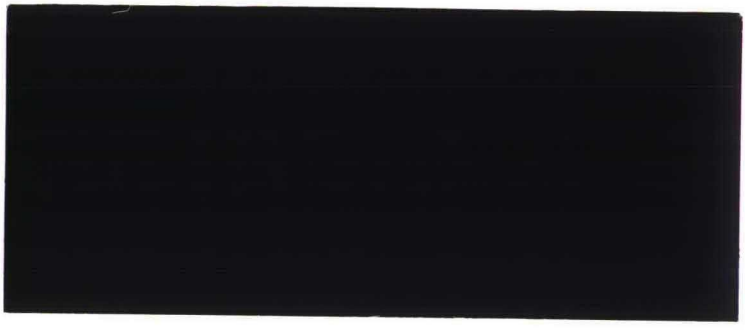
Regression sampling in statistical auditing

Kleijnen, J.P.C.; Kriens, J.; Timmermans, H.; van den Wildenberg, H.

Publication date:
1988



DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

REGRESSION SAMPLING IN STATISTICAL AUDITING

J.P.C. Kleijnen,
J. Kriens,
H. Timmermans and
H. Van den Wildenberg
FEW 306

January 1988

* J. Kleijnen is professor of Simulation and Information Systems in the Department of Information System and Auditing; J. Kriens is professor of Operations Research in the Department of Econometrics; H. Timmermans and H. Van den Wildenberg are graduate students in the School of Business and Economics of Tilburg University in Tilburg, the Netherlands.

Abstract

Different confidence intervals for the regression estimator are investigated. One interval is exact under the assumption of multivariate normal distributions. It gives longer intervals (hence better coverages) than the interval based on a popular variance estimator. An interval due to Roberts (1970) is much too long. Jackknifing gives robust intervals. The relationships with the difference and the ratio estimators are briefly discussed. A Monte Carlo experiment, based on the Neter and Loebbecke (1975) populations, gives estimated coverages and lengths of the different confidence intervals. Rules of thumb for practitioners are given.

CONTENTS

1. Introduction

Sampling has become well accepted in auditing. In particular, we consider the following practical situation (see Appendix 1 for a list of major symbols). There is a population of book values (say) y_i, and there is a corresponding population of actual (true) values x_i with $i = 1,\ldots,N$. For example, there are N stock keeping units (SKU's) which are shown in the books to have values y_i whereas their actual values are x_i. The auditor has to verify that the <u>total</u> book value $\sum_1^N y_i$ does not deviate "too much" from the total audit value $\sum_1^N x_i$ denoted by the symbol (say) ϑ. To save time, the auditor takes a sample of size n, without replacement, from the population (of size N), and observes both the true value $\underline{x}_j$ and the book value $\underline{y}_j$ (random variables are underlined), where $j = 1,\ldots,n$. This sample yields an estimator $\bar{\underline{x}} = \sum_1^n \underline{x}_j/n$ for the mean audit value (say) $\mu_x = \sum_1^N x_i/N = \vartheta/N$. The estimator $\bar{\underline{x}}$ is compared with the population mean $\sum_1^N y_i/N$. The total book value $\sum_1^N y_i$ is accepted as a substitute for the total (unknown) audit value $\sum_1^N x_i = \vartheta$, if the $1-\alpha$ confidence interval $[L_\alpha, U_\alpha]$ around the sample average $\bar{\underline{x}}$ "covers" the corresponding sample average $\bar{\underline{y}}$. In other words, by definition we have

$$1-\alpha = P(L_\alpha \leq \bar{\underline{x}} \leq U_\alpha) =$$

$$P(nL_\alpha \leq n\bar{\underline{x}} \leq nU_\alpha) \qquad (1.1)$$

So if $L_\alpha < \sum_1^N y_i/N < U_\alpha$ then the auditor does not reject the hypothesis

$$H_0 : \vartheta(\equiv \sum_1^N x_i) = \sum_1^N y_i. \tag{1.2}$$

The problem investigated in this paper is the construction of a confidence interval that has indeed the correct probability 1-α. If the $\underline{x}_j$ are Normally and Independently Distributed with (constant) mean μ_x and (constant) variance σ_x^2 - or $\underline{x}_j \sim \text{NID}(\mu_x, \sigma_x^2)$ - then elementary mathematical statistics yields a 1-α confidence interval (see eq. 2.1). If, however, the $\underline{x}_j$ are skewly distributed (as they are in auditing; see table 2), then the classical interval has a coverage probability smaller than 1-α; see Kleijnen (1987, pp. 18-23). Moreover (whether the $\underline{x}_j$ are normally or not normally distributed) the confidence interval may be too long for practical purposes (see § 2). Then the 1-α confidence interval for the audit values can be made much shorter by using information on the book values. This information on the auxiliary variable (book values) leads to several estimators: ratio, difference, and regression (or control) estimators. This paper focusses on regression estimators, but does include brief references to the other two estimators; see § 7. We shall study the following four 1-α confidence intervals for the regression estimator, plus one classical interval:

(i) Use the normal distribution and a popular variance estimator, as Kriens and Peterse (1986) did; see § 3.

(ii) Use the Student distribution and the Ordinary Least Squares variance estimator which gives an optimal 1-α confidence interval provided $\underline{x}_j$ and $\underline{y}_j$ are bivariate-normal; see Lavenberg et al. (1982); see § 5.

(iii) Use the Student distribution and jackknifing. Jackknifing is a very general "trick" - see Miller (1974) - which was applied to the ratio estimator by Frost and Tamura (1982); see § 6.

(iv) Roberts (1978) proposed a special confidence interval procedure for the regression estimator in auditing; see § 4.

(v) We also study the classical or crude procedure based on $\underline{x}_j \sim NID(\mu_x,\sigma_x^2)$; see § 2.

For these five approaches we use Monte Carlo simulation, based on the Neter and Loebbecke (1972) auditing populations (see § 9), in order to estimate the (actual) coverages of the confidence intervals (with nominal coverages of $1-\alpha$), and their lengths; see § 10. Our study yields guidelines for practitioners; see § 11.

2. The classical or crude estimator $\bar{\underline{x}}$

Elementary mathematical statistics proves that

$$P(\bar{\underline{x}} - t_{n-1}^{(\alpha/2)}(1-\frac{n}{N})^{\frac{1}{2}}\frac{s_{\underline{x}}}{\sqrt{n}} \leq \mu_x \leq \bar{\underline{x}} + t_{n-1}^{(\alpha/2)}(1-\frac{n}{N})^{\frac{1}{2}}\frac{s_{\underline{x}}}{\sqrt{n}}) = 1-\alpha \quad (2.1.a)$$

where $t_{n-1}^{(\alpha/2)}$ is the upper $1-(\alpha/2)$ "quantile" of the Student statistic with $n-1$ degrees of freedom, that is

$$P(\underline{t}_{n-1} \leq t_{n-1}^{(\alpha)}) = 1-\alpha \quad (2.1.b)$$

and $\bar{\underline{x}}$ is the (sample) average or crude estimator

$$\bar{\underline{x}} = \frac{\sum_{1}^{n} \underline{x}_j}{n} \qquad (2.1.c)$$

and $\underline{s}_x$ is the estimated standard deviation, that is, it is the square root of the variance estimator

$$\underline{s}_x^2 = \frac{\sum_{1}^{n} (\underline{x}_j - \bar{\underline{x}})^2}{n-1} = \frac{\sum_{1}^{n} \underline{x}_j^2}{n-1} - \frac{n}{(n-1)} \bar{\underline{x}}^2 \qquad (2.1.d)$$

and $(1 - \frac{n}{N})^{\frac{1}{2}}$ is the finite-population correction (which approaches one as N approaches infinity).

As we mentioned in § 1, if the total book value $\sum_{1}^{N} y_i$ lies within the $1-\alpha$ confidence interval (2.1.a), then the auditor accepts the total book value $\sum_{1}^{N} y_i$ as representing the unknown true value $\vartheta (= \sum_{1}^{N} x_i = N\mu_x)$. Such a procedure involves the risk of erroneously accepting $\sum_{1}^{N} y_i$: so-called type II error. This risk increases as the confidence interval's length increases. To decrease this risk, auditors may work with estimators more refined than the "crude" estimator $\bar{\underline{x}}$.[1)]

3. The regression estimator $\bar{\underline{x}}(\hat{\underline{y}})$

Estimators more efficient than the crude estimator $\bar{\underline{x}} (= \sum_{1}^{n} \underline{x}_j / n)$, also observe the corresponding book values in the sample, namely $\underline{y}_j$. The

difference estimator or control-variate estimator corrects the crude estimator $\bar{\underline{x}}$ for the deviation between the estimator $\bar{\underline{y}}$ and the known population value $\mu_y = \sum_1^N y_i/N$:

$$\bar{\underline{x}}(\gamma) = \bar{\underline{x}} + \gamma(\mu_y - \bar{\underline{y}}). \tag{3.1}$$

Obviously, for any given constant γ, the estimator $\bar{\underline{x}}(\gamma)$ is an unbiased estimator of μ_x.

It is well-known (and simple to prove) that the variance of $\bar{\underline{x}}(\gamma)$ is minimized by $\gamma_0 = \text{cov}(\underline{x},\underline{y})/\text{var}(\underline{y})$. The obvious estimator for γ_0 is

$$\hat{\underline{\gamma}} = \frac{\hat{\text{cov}}(\underline{x},\underline{y})}{\hat{\text{var}}(y)} = \frac{\hat{\rho}(\underline{x},\underline{y})\underline{s}_x}{\underline{s}_y}, \tag{3.2}$$

where $\hat{\rho}$ denotes the estimated correlation. This $\hat{\underline{\gamma}}$ is identical to the Ordinary Least Squares (OLS) estimator $\hat{\underline{\gamma}}$ in

$$\hat{\underline{x}} = \hat{\underline{\delta}} + \hat{\underline{\gamma}}\,\underline{y}, \tag{3.3}$$

in other words, the estimator $\hat{\underline{\gamma}}$ can be based on a familiar numerical algorithm, available on any computer. Substitution of the estimator $\hat{\underline{\gamma}}$ into (3.1) results in the so-called regression estimator, a non-linear estimator:

$$\bar{x}(\hat{\underline{\gamma}}) = \bar{\underline{x}} + \hat{\underline{\gamma}}(\mu_y - \bar{\underline{y}}). \tag{3.4}$$

This estimator is no longer unbiased. Its variance can be estimated in several ways; Deng and Wu (1987) prove that many popular variance estimators underestimate the true variance. Kriens and Peterse (1986, p. 28) use:

$$s^2\{\bar{x}(\hat{\bar{y}})\} = \frac{(1 - \frac{n}{N})}{n(n-1)} \left[\sum_1^n x_j^2 - n\bar{x}^2 - \frac{(\sum_1^n x_j y_j - n\bar{x}\bar{y})^2}{\sum_1^n y_j^2 - n\bar{y}^2} \right]. \qquad (3.5)$$

In § 5 we shall see that this estimator is biased; it underestimates the variance. If $\underline{s}$ denotes $[s^2\{\bar{x}(\hat{\bar{y}})\}]^{\frac{1}{2}}$ and $z^{(\alpha/2)}$ is the $1-(\alpha/2)$ quantile of the standard normal $\underline{z}$, then the $1-\alpha$ confidence interval is

$$[\bar{x}(\hat{\bar{y}}) - z^{(\alpha/2)}\underline{s},\ \bar{x}(\hat{\bar{y}}) + z^{(\alpha/2)}\underline{s}]. \qquad (3.6)$$

They conjecture - also see Baker and Copeland (1979) - that this confidence interval gives correct coverage $(1-\alpha)$, if the following conditions hold.

(i) The sample size n is at least 200.

(ii) The number of non-zero differences $(x_j \neq y_j)$, say $\underline{m}$, is at least 5% of n, and at least 20.

(iii) There are both positive and negative differences, to be tested by the sign statistic. (The sign test concerns H_0 : $p = 0.50$ where $p = P(\underline{x}-\underline{y}>0) = P(\underline{x}-\underline{y}\leq 0)$.)

Our study examines these conditions, as we shall see. Here we mention that condition (ii) is the central issue of the following procedure.

4. The Roberts (1978) procedure

Roberts (1978, pp. 90-91) proposes a confidence interval for the regression estimator $\bar{x}(\hat{\gamma})$ which accounts for $\underline{m}$, the number of non-zero differences. Suppose the population has a low error percentage, that is, most book values y_i are identical to the corresponding audited values x_i with $i = 1,\ldots,N$. Then $\underline{m}$ is small, with high probability. Roberts' procedure accounts for $\underline{m}$. Unfortunately, his procedure is lengthy and difficult to understand intuitively. Because our Monte Carlo experiment shows that his procedure is completely inferior, we do not present his formulas in the main text but move them to Appendix 2. Further, our Monte Carlo study gave extremely long confidence intervals for Roberts' procedure; so we changed the first step of his procedure (his factor $P_U(m)$ is an upper bound for $\underline{m}/n$; we replace $P_U(m)$ by m/n; for example, for $\underline{m} = 15$ and $n = 180$ Roberts gives $P_U(\underline{m}) = 0.13$ whereas we use $\underline{m}/n = 0.08$, so that in step 5 of the procedure a lower estimated standard deviation follows; also see Appendix 2).

5. The Lavenberg et al. (1982) bivariate-normal model

Lavenberg et al. (1982) discuss situations with $Q \geq 1$ auxiliary or control variables (instead of a single control variate as in our auditing situation). In Appendix 3 we quote their text, and we give a translation into our symbols. However, it is simpler to apply their idea immediately to the simple regression model (3.3).

Eqs. (3.2) and (3.3) showed that we compute the estimate $\hat{\gamma}$ for the regression estimator $\bar{x}(\hat{\gamma})$, using OLS. (We note that the OLS criterion

results in a purely mathematical algorithm.) Now we replace the linear model (3.3) by

$$E(\underline{x}|y) = \delta + \gamma(y-\mu_y) \tag{5.1}$$

and assume that the distribution $F(\underline{x}|y)$ is normal with (constant) variance σ^2 (in other words, the variance does not depend on y). Actually, it can be proved that this model holds, if (but not only if) the pair $(\underline{x},\underline{y})$ is bivariate-normally distributed. Then the familiar OLS formulas give the Best Linear Unbiased Estimator (BLUE) of the regression parameters δ and γ in (5.1):

$$\hat{\underline{\gamma}}' = \frac{\sum_1^n (y_j-\bar{y})(\underline{x}_j-\bar{\underline{x}})}{\sum_1^n (y_j-\bar{y})^2} \tag{5.2}$$

where the dash distinguishes this conditional estimator from the unconditioned estimator $\hat{\underline{\gamma}}$ in (3.2); further we have:

$$\hat{\underline{\delta}}' = \bar{\underline{x}} + \hat{\underline{\gamma}}'(\mu_y-\bar{y}) \equiv \bar{x}(\hat{\underline{\gamma}}|y). \tag{5.3}$$

We also know that $\hat{\underline{\Omega}}$, the estimator for the covariance matrix of the estimated regression parameters, is:

$$\hat{\underline{\Omega}} = (W'W)^{-1}\,\hat{\underline{\sigma}}^2, \tag{5.4}$$

where W is the n×2 matrix of explanatory variables corresponding to (5.1):

$$W = \begin{bmatrix} 1, & y_1 - \mu_y \\ \vdots & \\ 1, & y_n - \mu_y \end{bmatrix} \tag{5.5}$$

and $\hat{\underline{\sigma}}^2$ is the Mean Squared Residual:

$$\hat{\underline{\sigma}}^2 = \frac{\sum_1^n [\underline{x}_j - \{\hat{\underline{\delta}}' + \hat{\underline{\gamma}}'(y_j - \mu_y)\}]^2}{n-2} = \frac{\Sigma[(\underline{x}_j - \bar{\underline{x}}) - \hat{\underline{\gamma}}'(y_j - \bar{y})]^2}{n-2}. \tag{5.6}$$

Let v_{11} denote the element (1,1) of $(W'W)^{-1}$; see (5.4). Then we have for $\delta' = E(\hat{\underline{\delta}}') = E[\bar{x}(\hat{\underline{\gamma}}|y)]$:

$$P[\hat{\underline{\delta}}' - t_{n-2}^{(\alpha/2)}(v_{11}\hat{\underline{\sigma}}^2)^{\frac{1}{2}} \leq \delta' \leq \hat{\underline{\delta}}' + t_{n-2}^{(\alpha/2)}(v_{11}\hat{\underline{\sigma}}^2)^{\frac{1}{2}}] = 1-\alpha. \tag{5.7}$$

Since the right-hand side does not depend on y, it also holds unconditionally, that is, in the above equations we can make y random, as is the case for the regression estimator $\bar{x}(\hat{\underline{\gamma}})$. So the point estimator of $\mu_x = \vartheta/N$ is the OLS estimator $\hat{\underline{\delta}}$ and its confidence interval follows from (5.7) where we now make the auxiliary variable stochastic ($\underline{y}$), so that $\hat{\underline{\gamma}}'$ and $\hat{\underline{\delta}}'$ become $\hat{\underline{\gamma}}$ and $\hat{\underline{\delta}} \equiv \bar{x}(\hat{\underline{\gamma}})$, and v_{11} becomes $\underline{v}_{11}$.

It is easily proved that

$$\underset{\sim}{v}_{11}\,\hat{\underset{\sim}{\sigma}}^2 = \left[\frac{1}{n} + \frac{(\bar{\underset{\sim}{y}}-\mu_y)^2}{\sum_1^n (\underset{\sim}{y}_j-\bar{\underset{\sim}{y}})^2}\right] \times \left[\Sigma(\underset{\sim}{x}_j-\bar{\underset{\sim}{x}})^2 - \frac{\{\Sigma\ (\underset{\sim}{x}_j-\bar{\underset{\sim}{x}})(\underset{\sim}{y}_j-\bar{\underset{\sim}{y}})\}^2}{\Sigma\ (\underset{\sim}{y}_j-\bar{\underset{\sim}{y}})^2}\right] \times$$

$$\times\ (n-2)^{-1} \qquad (5.8)$$

which may be compared to the variance estimator used by Kriens and Peterse (1986); see eq. (3.5). This shows the following differences between the confidence intervals of eqs. (3.6) and (5.7):

(i) Lavenberg et al. (1982) do not use the finite-population correction $(1-n/N)^{\frac{1}{2}}$ to compute the standard error. In our Monte Carlo experiment this factor ranges between 0.975 (n = 200; N = 4033) and 0.994 (n = 100, N = 8309). This factor decreases Kriens and Peterse's variance estimator and hence the coverage.

(ii) Lavenberg et al. use the denominator n-2, that is, the total number of sample observations corrected for the number of regression parameters to be estimated. Kriens and Peterse use the denominator n-1, that is, they correct only for the estimation of the mean μ_x. This factor slightly decreases Kriens and Peterse's variance estimator: $\{(n-1)/(n-2)\}^{\frac{1}{2}}$ = 1.005 and 1.003 for n = 100 and 200 respectively. Actually (by accident) we programmed Kriens and Peterse's estimator using n-2 instead of n-1.

(iii) Lavenberg et al. use the factor $(\bar{\underset{\sim}{y}}-\mu_y)^2/\Sigma(\underset{\sim}{y}_j-\bar{\underset{\sim}{y}})^2$ whose value depends on the sample; anyhow this factor is positive (albeit small) and increases the estimated variance. (Note that $E[(\bar{\underset{\sim}{y}}-\mu_y)^2/\Sigma(\underset{\sim}{y}_j-\bar{\underset{\sim}{y}})^2] \approx E[(\bar{\underset{\sim}{y}}-\mu_y)^2]/E[\Sigma(\underset{\sim}{y}_j-\bar{\underset{\sim}{y}})^2]$ = $\mathrm{var}(\bar{\underset{\sim}{y}})/\{\mathrm{var}(\underset{\sim}{y})(n-1)\}$ = $1/\{n(n-1)\}$ which approaches zero for large n. If $\underset{\sim}{y}$ is normally distributed then

$(\bar{\underline{y}}-\mu_y)^2/\sum_1^n (\underline{y}_j-\bar{\underline{y}})^2 = \frac{1}{n^2}\underline{F}_{1,n-1}$ so that its mean - see Wilks (1962, p. 187) - is $(n-1)/\{(n-3)n^2\}$.

(iv) Lavenberg et al. use $t_{n-2}^{(\alpha/2)}$ in (5.7) whereas Kriens and Peterse use $z^{(\alpha/2)}$ in (3.7). For $\alpha = 0.05$ we use $z^{(\alpha/2)} = 1.96$, $t_{98}^{(\alpha/2)} = 1.9845$ and $t_{198}^{(\alpha/2)} = 1.972$ (the t values are computed by a SAS subroutine).

(v) Numerical inaccuracies differ if different computational procedures are used. For example mathematically we have the identity: $\Sigma(x_j-\bar{x})^2 = \Sigma x_j^2 - n\bar{x}^2$. The second expression, however, gives a smaller numerical accuracy. We programmed the equations exactly as formulated by the authors.

Each of the first three factors increases Lavenberg et al.'s variance estimator. The fourth factor also increases the confidence interval's length and hence the coverage.

We note that the bivariate-normal model, which implies constant variances, may be violated in auditing applications; see Beck (1980), Frost and Tamura (1982) and § 9.1.

6. Jackknife

We first explain jackknifing in general; next we apply the jackknife to the regression estimator. Suppose we have a sample of size n for the random variables $\underline{z}_j$ (j = 1,...,n). This sample yields the estimator (say) $\hat{\underline{\theta}}$; for example, $\hat{\underline{\theta}}$ denotes the sample average $\bar{\underline{z}}$. Next we delete variable j and from the (n-1) remaining variables we compute the same estimator $\hat{\underline{\theta}}_{-j}$; for example, $\bar{\underline{z}}_{-1} = \sum_{j_1'=2}^{n} \underline{z}_{j_1'}/(n-1)$. Obviously we have n estimators

$\hat{\underline{\theta}}_{-1},\ldots,\hat{\underline{\theta}}_{-n}$; for example $\bar{\underline{z}}_{-1},\ldots,\bar{\underline{z}}_{-n}$ $(= \sum_{j'_n=1}^{n-1} \underline{z}_{j'_n}/(n-1))$. Now we define the n pseudo-values (say) $\underline{J}_j$:

$$\underline{J}_j = n\hat{\underline{\theta}} - (n-1)\hat{\theta}_{-j} \qquad (j = 1,\ldots,n). \tag{6.1}$$

It can be proved that $\underline{J}_j$ has smaller bias than $\hat{\underline{\theta}}$ possibly had. Moreover, even if the variables $\underline{z}_j$ are not normally distributed, the jackknife confidence interval "often" holds (the interval is "robust"). This interval is computed straightforwardly, treating the $\underline{J}_j$ as if they were normally and independently distributed (also see eq. 2.1):

$$1-\alpha \approx P\left(\frac{|\bar{\underline{J}} - \theta|}{\underline{s}_J/\sqrt{n}} \leq t_{n-1}^{(\alpha/2)}\right) \tag{6.2}$$

where $\bar{\underline{J}} = \sum_1^n \underline{J}_j/n$ and $\underline{s}_J^2 = \sum_1^n (\underline{J}_j - \bar{\underline{J}})^2/(n-1)$. See Miller (1974) and also Kleijnen (1987, p. 424).

Jackknifing the regression estimator is now very simple. We had a sample of n (independent) pairs $(\underline{x}_j, \underline{y}_j)$ from which we computed the regression estimator $\bar{x}(\hat{\underline{\gamma}})$; see (3.4). Now we delete pair j (j = 1,...,n) and from the remaining (n-1) pairs we again compute the regression estimator (say) $\bar{x}(\hat{\underline{\gamma}}_{-j})$; see Appendix 4. This yields the pseudo-values; see (6.1) with $\hat{\underline{\theta}} = \bar{x}(\hat{\underline{\gamma}})$ and $\hat{\underline{\theta}}_{-j} = \bar{x}(\hat{\underline{\gamma}}_{-j})$. Finally (6.2) yields the 1-α confidence interval for μ_x. Lavenberg et al. (1982) also examined the jackknifing regression estimator, but not in an auditing situation.

Deng and Wu (1987) prove that the jackknifed regression estimator yields an overestimated variance: $E(\underline{s}_J^2)$ is too high (they did not use an auditing context). Frost and Tamura (1982) applied jackknifing to a related estimator, namely the ratio estimator; also see the next section.

7. Related estimators: difference and ratio estimators

If the book values are identical to the audited values, then $\hat{\underline{\gamma}} = 1$ and $\hat{\underline{\delta}} = 0$ in (3.3) and the regression estimator reduces to the <u>difference estimator</u> with coefficient one (see eq. 3.1):

$$\bar{\underline{x}}(1) = \bar{\underline{x}} + (\mu_y - \bar{\underline{y}}) = \frac{\sum_{1}^{n} (\underline{x}_j - \underline{y}_j)}{n} + \mu_y = 0 + \mu_y = \mu_y. \qquad (7.1)$$

Lavenberg et al. (1982) point out that estimation of the (optimal) γ in (3.1) reduces the variance reduction achieved by the regression estimator. Therefore we might argue that it is better not to estimate the exact optimal value, but to substitute $\gamma = 1$ into (3.1). However, if $\hat{\delta} = 0$ (and $\hat{\gamma} = 1$) then it seems better to use another related estimator, namely the <u>ratio estimator</u> (say) $\bar{\underline{x}}_r$:

$$\bar{\underline{x}}_r = \bar{\underline{x}} \frac{\mu_y}{\bar{\underline{y}}}. \qquad (7.2)$$

Deng and Wu (1987, p. 68) state: "The ratio estimator is appropriate for populations whose regression line passes close to the origin"; also see

Beck (1980, p. 16). We further note that computation of a regression estimator with $\hat{\delta} = 0$ (and arbitrary $\hat{\gamma}$) would be simple, because most software makes it possible to force the regression equation to pass through the origin.

If indeed the book values are (nearly) identical to the audited values, then we do not need anything more efficient than the crude estimator $\bar{x}$. That estimator has the correct α error (assuming normality); even though its confidence interval may be wide, we cannot make a β error since the assumption was that the book values are correct. If the book values may be fraudulent, then we need a more efficient estimator. Then, however, it is not sure that $\hat{\gamma} = 1$ or $\hat{\delta} = 0$ in (3.3), in other words, we conjecture that the difference and the ratio estimator are less efficient in situations where efficiency is most important!

Instead of relying on this prior knowledge about auditing applications, we might reason as follows. The regression estimator (including the difference estimator) corrects for <u>absolute</u> deviations $(\mu_y - \bar{y})$, whereas the ratio estimator corrects for <u>relative</u> deviations $\mu_y/\bar{y}$. Which estimator is better? We might compute the estimated correlation coefficient between x and y (say) $\hat{\rho}_1 = \hat{\rho}(x,y)$ and between $\ell n\, \bar{x}$ and $\ell n\, \bar{y}$ (say) $\hat{\rho}_2 = \hat{\rho}(\ell n\, x, \ell n\, y)$. If $\hat{\rho}_1 < \hat{\rho}_2$ then we use the ratio estimator. (We might even use both estimators and, applying the Bonferroni inequality, derive a single confidence interval.) In our study, however, we investigate the regression estimator only.

We could have refined the regression estimator by introducing Weighted Least Squares (WLS) to estimate γ. WLS, however, is more complicated than OLS, especially since we have to estimate the heterogeneous variances. Beck (1980, pp. 19, 25) states that WLS and OLS are "very similar". Variances proportional to the book values are also discussed in Frost and Tamura (1982).

8. Intermezzo: a half-way summary

To summarize so far, the crude estimator $\bar{\underline{x}}$ may yield too wide a $1-\alpha$ confidence interval, so that the risk of (erroneously) accepting a false book value ($N\mu_y$) is too high (see § 2). The regression estimator $\bar{\underline{x}}(\hat{\gamma})$ has a smaller variance, if $\hat{\gamma}$ is well chosen. The optimal $\hat{\gamma}$ can be estimated by the OLS algorithm. Unfortunately, an exact confidence interval is unknown (§ 3), unless the pairs $(\underline{x}_j, \underline{y}_j)$ are bivariate-normal (§ 5).

Note that if the $\underline{x}_j$ are not normal, then the crude estimator $\bar{\underline{x}}$ does not have an exact $1-\alpha$ confidence interval either. Kleijnen et al. (1986) found good coverage, if the crude estimator $\bar{\underline{x}}$ is combined with a t statistic modified such that the skewness of the distribution of $\underline{x}$ is estimated.

Robust intervals can be based on jackknifing (§ 6). We also mentioned Roberts' procedure in case $\underline{m}$ (the number of non-zero differences $\underline{x}_j - \underline{y}_j$) is small (§ 4).

9. Monte Carlo input or experimental design

In this section we shall first describe the populations from which we sample in our Monte Carlo experiment (see § 9.1). Next we shall describe some other inputs to the Monte Carlo experiment (see § 9.2).

9.1. The Neter and Loebbecke (1975) populations

We use a set of realistic data developed by Neter and Loebbecke (1975, pp. 11-32), also used in several other studies on statistical auditing.[2] The data are characterized by Frost and Tamura (1982, pp. 109-111) as follows (square brackets denote our comments; three dots indicate that we deleted some text in the original):

> "...Their data set is based on four real-world accounting populations with a variety of error characteristics. For each accounting population five study populations with different error rates are developed, yielding a total of twenty study populations [an error rate of 30% means that 0.30N items are wrongly booked]...
> Some of the key characteristics of the error distribution are reported in table 1. Except for accounting population 2, the error rates are 30, 10, 5, 1 and ½ percent. Accounting population 2 differs in that it has a 70-percent error rate instead of the 30-percent rate. Column 3 of table 1 reports the total audit value in thousands of dollars [θ]. Columns 4 through 7 show the mean, standard deviation, skewness, and kurtosis of the errors. Accounting populations 1 and 2 have both positive and negative

Table 1*

Characteristics of Error Distribution in Study Populations

Accounting Population	Error Rate	$[\vartheta]$ ($1.000)	Distribution of Errors				Heteroscedas.		cv(y) n = 100
			Mean	σ	Skewness	Kurtosis	$\hat{\gamma}$	$s(\hat{\gamma})$	
1	30%	380	-.09	4.1	-1.1	72.6	.86	.19	.291
1	10%	380	-.05	2.7	-1.8	186.1	.96	.23	.291
1	5%	379	.01	1.9	6.2	466.1	.90	.32	.291
1	1%	379	-.02	.9	-30.9	1442.6	.94	.47	.291
1	½%	379	-.01	.5	-31.0	2287.3	-.40	.53	.291
2	70%	3.565	-14.2	194.3	-1.5	41.3	.94	.08	.182
2	10%	3.491	-.8	80.9	-1.4	360.2	1.11	.12	.182
2	5%	3.491	-.8	63.9	-7.4	748.8	1.48	.16	.182
2	1%	3.486	.2	25.7	48.4	3267.7	.97	.37	.182
2	½%	3.487	-.1	8.6	-9.7	912.0	.22	.34	.182
3	30%	13.510	23.0	72.7	4.5	22.4	.28	.31	.361
3	10%	13.623	6.9	39.6	8.4	82.4	.20	.30	.361
3	5%	13.648	3.4	27.0	12.1	176.1	.26	.30	.361
3	1%	13.666	.8	13.5	26.5	827.0	-.01	.36	.361
3	½%	13.669	.4	9.7	42.0	2029.4	-.30	.38	.361
4	30%	6.442	263.0	1529.7	9.9	113.8	1.92	.02	.208
4	10%	7.237	65.9	662.9	19.6	471.5	1.83	.04	.208
4	5%	7.402	24.9	302.5	25.3	840.3	1.74	.06	.208
4	1%	7.469	8.5	222.9	46.1	2448.6	1.60	.40	.208
4	½%	7.478	6.2	216.5	49.9	2751.5	1.44	.41	.208

* Reproduced from Frost and Tamura (1982, p. 110).

errors. The mean error is close to zero for accounting population 1. Since all of the errors in accounting populations 3 and 4 are overstatements, the mean errors for these populations are substantially greater than zero. The error distributions in all of the study populations exhibit skewness and kurtosis. The extreme nonnormality of the error distributions is clear from these parameters. By using the ratio of the standard deviation of the error to the total audit value as a measure of the size of the error, one can see that accounting population 4 has relatively large errors and accounting populations 1 and 3 have relatively small errors.

An important characteristic of the error distribution is the relationship between its dispersion and the book value (i.e., the degree of heteroscedasticity)...

Except for the ½-percent error rates, $\hat{\gamma}$ is close to one in accounting populations 1 and 2, indicating that the variance of the errors is approximately proportional to the book value... In accounting population 3, $\hat{\gamma}$ is close to zero, suggesting that the errors are homoscedastic in this population. In accounting population 4, $\hat{\gamma}$ is greater than one, implying that the variance of the errors increases more than proportionally to book value.

The last column reports the coefficient of variation of $\bar{y}$ for a sample size of 100..."

Table 2*

Neter-Loebbecke Populations of Account Book Values

Population	1	2	3	4
Total book value	$379,131	$3,486,530	$13,671,500	$7,502,957
Mean	45.63	636	1,946	1,860
Standard deviation	132.61	1,156	7,022	3,865
Skewness	22.0	3.5	7.9	3.2
Kurtosis	906.4	15.2	78.1	11.4
Largest account	6,869.70	9,989	98,163	24,928.6
Smallest account	.50	1.00	.10	.10
Error pattern	small, offsetting	moderate, primarily understatement	small overstatement	large overstatement (heteroscedastic)

* Reproduced from Beck (1980, p. 23).

Beck (1980) characterizes, not the errors $\underline{x} - \underline{y}$, but the book values $\underline{y}$; see table 2 and Neter and Loebbecke (1975). In the Monte Carlo experiment different error rates create different audit values x_i, not different book values y_i ($i = 1,\ldots,N$).

Neter and Loebbecke (1975) created some more populations by modifying accounting population 1; that is, book values higher than $ 950 are not sampled but are investigated individually. We utilized such a modified population, denoted by 1M (see the tables in the next section, § 10).

Obviously, the distribution of $\underline{y}$ is not Gaussian, especially populations 1 and 1M are not (see table 2). We emphasize that a logarithmic transformation of $\underline{x}$ and $\underline{y}$ does not help; that is, $\log(\underline{x})$ and $\log(\underline{y})$ are more normally distributed, but the $1-\alpha$ confidence interval for $E\{\log(\underline{x})\}$ does not yield a $1-\alpha$ confidence interval for $E(\underline{x}) = \vartheta'$; see Patterson (1966).

9.2. Other Monte Carlo inputs

We decided to study sample sizes equal to 200 (the minimum size required by Kriens and Peterse, 1986) and 100. (Frost and Tamura, 1982, study n = 50, 100, 200; Beck, 1980, p. 20, took n = 200, 600.)

Though the auditor samples without replacement, we sample with replacement, because the latter procedure is simpler; both procedures are equivalent from a practical viewpoint, the factor n/N being smaller than

0.05. We repeat this sampling 600 times (as Neter and Loebbecke, 1975, also did; Beck (1980, p. 20) took only 300 replications).

The pseudorandom numbers are created by a multiplicative congruential generator with multiplier 13^{13} and modulo 2^{59}, developed by NAG (Numerical Algorithms Group, United Kingdom). We never reset the seed, so that all results in the following tables are independent, except for results on the same line (the latter results are obtained by applying different estimators to the same sample data $\underline{x}_j$, $\underline{y}_j$).

10. Monte Carlo output

We are interested in the question whether the constructed "nominal" 1-α confidence interval indeed covers the true mean ϑ $(= \sum_1^N x_i)$? The Roberts (1978) procedure requires a number of non-zero differences higher than two ($\underline{m}>2$) (he uses Student's t with m-2 degrees of freedom). Therefore we do not compute a confidence interval if $\underline{m} \leq 2$. Let $\underline{K}$ denote the number of iterations with $\underline{m} > 2$ or $\underline{m} = 0$: if $\underline{m} = 0$ then we do not reject H_0. (obviously $0 \leq \underline{K} \leq 600$; see § 9.2). Define

$\underline{v}_k = 0$, if the (1-α) confidence interval does not cover the true mean;

$= 1$, else $(k = 1,\ldots,\underline{K})$. (10.1)

Then the estimator for the coverage equals $\sum_{1}^{K} v_k/K$. A perfect 1-α confidence interval has expected coverage equal to 1-α. A "conservative" interval has expected coverage higher than 1-α (the interval from zero to infinity is certainly conservative). Therefore we also estimate the expected length of the nominal (1-α) confidence interval (we again estimate this length from K iterations).

We emphasize that we also compute a confidence interval in case (nearly) all differences $x_j - y_j$ have the same sign. If in practice one does not compute a confidence interval in this case, then the coverages will be higher than our estimated coverages.

Table 3 indicates that Roberts (1978) gives too high a coverage; Table 5 shows that such a coverage implies a very long confidence interval. An exception is population 4 (large, overstatement; see table 2). However, for population 4 the crude estimator behaves somewhat better than Roberts' estimator; its confidence interval length remains ridiculous (Table 5). Kriens and Peterse (1986) gives results very close to Lavenberg et al. (1982), as we know (see § 5). Jackknifing gives coverages which in general are closest to the nominal value 1-α (= 0.95); consequently its confidence intervals are slightly longer than the intervals based on Kriens and Peterse (1986) and Lavenberg et al. (1982); see tables 3 and 5.

Table 3

Estimated Coverages for Nominal $1-\alpha = 0.95$ and Sample Size $n = 100$

Population	Error Rate (%)	Crude Est.	Kriens Peterse	Lavenberg et al.	Jackknife	Roberts (1978)
1	0.5	.82	1.00	1.00	1.00	1.00
1	1.0	.84	.82	.83	.83	1.00
1	5.0	.79	.95	.96	.99	1.00
1	10.0	.79	.89	.90	.93	1.00
1	30.0	.80	.87	.89	.94	1.00
1M	1.0	.87	.84	.86	.84	1.00
1M	10.0	.88	.90	.91	.94	1.00
2	0.5	1.00	1.00	1.00	1.00	1.00
2	1.0	.94	.96	.96	.96	1.00
2	5.0	.91	.85	.86	.87	.99
2	10.0	.90	.97	.98	.98	1.00
2	70.0	.92	.93	.93	.97	1.00
3	0.5	.74	1.00	1.00	1.00	1.00
3	1.0	.82	.91	.92	.92	1.00
3	5.0	.83	.77	.78	.78	.94
3	10.0	.83	.78	.79	.79	.96
3	30.0	.83	.89	.90	.90	1.00
4	0.5	.92	.75	.75	.75	.83
4	1.0	.86	.86	.87	.86	.98
4	5.0	.93	.59	.60	.60	.88
4	10.0	.93	.60	.61	.62	.84
4	30.0	.91	.71	.71	.75	.97

Table 4

Estimated Coverages for Nominal $1-\alpha = 0.95$ and Sample Size $n = 200$

Population	Error Rate (%)	Crude Est.	Kriens Peterse	Lavenberg et al.	Jackknife	Roberts (1978)
1	0.5	.85	.84	1.00	.84	1.00
1	1.0	.87	.76	.76	.78	1.00
1	5.0	.84	.95	.96	.98	1.00
1	10.0	.82	.91	.92	.95	1.00
1	30.0	.85	.89	.89	.95	1.00
1M	1.0	.89	.80	.81	.83	1.00
1M	10.0	.89	.90	.91	.94	1.00
2	0.5	.90	.96	.96	.96	1.00
2	1.0	.94	.84	.84	.85	1.00
2	5.0	.94	.91	.91	.92	.99
2	10.0	.95	.96	.96	.97	1.00
2	70.0	.94	.94	.95	.96	1.00
3	0.5	.90	.92	.92	.92	.96
3	1.0	.88	.87	.87	.87	.98
3	5.0	.87	.81	.82	.81	.97
3	10.0	.88	.88	.89	.89	1.00
3	30.0	.87	.91	.92	.91	1.00
4	0.5	.91	.83	.83	.83	.95
4	1.0	.92	.74	.74	.73	.99
4	5.0	.93	.66	.66	.66	.90
4	10.0	.94	.68	.69	.69	.95
4	30.0	.93	.81	.82	.84	1.00

Table 5

Average Lengths of Estimated Confidence Intervals for Sample Size n = 100 and Nominal 1-α = 0.95

Population	Error Rate(%)	Crude Est.	Kriens Peterse	Lavenberg et al.	Jackknife	Roberts (1978)
1	0.5	385,014	5,216	5,382	5,171	47,793
1	1.0	344,054	3,103	3,192	3,258	29,583
1	5.0	322,973	4,489	4,649	5,073	20,973
1	10.0	323,408	6,359	6,566	7,146	25,720
1	30.0	316,749	11,157	11,518	13,666	69,533
1M	1.0	220,380	2,506	2,583	2,500	25,443
1M	10.0	221,487	6,520	6,694	7,018	26,814
2	0.5	2,609,349	27,045	27,820	27,971	293,793
2	1.0	2,407,226	64,187	65,876	67,011	538,035
2	5.0	2,421,411	92,994	95,691	99,053	402,242
2	10.0	2,372,514	134,004	137,886	143,257	519,596
2	70.0	2,246,366	385,345	393,153	419,726	3,349,842
3	0.5	15,903,616	74,486	78,292	73,780	796,787
3	1.0	16,538,144	51,580	53,305	51,697	469,808
3	5.0	17,063,743	61,549	63,945	62,816	275,057
3	10.0	17,187,754	95,804	99,289	97,183	372,218
3	30.0	17,399,247	190,494	197,935	194,150	1,141,977
4	0.5	6,269,035	616,578	637,667	690,923	4,605,579
4	1.0	6,014,548	292,392	301,264	323,596	2,590,072
4	5.0	5,947,289	301,399	310,874	321,243	1,306,091
4	10.0	5,861,363	659,940	680,019	741,663	2,658,419
4	30.0	5,483,599	1,751,350	1,806,252	2,095,786	11,340,530

Table 6

Average Lengths of Estimated Confidence Intervals for Sample Size n = 200 and Nominal 1-α = 0.95

Population	Error Rate(%)	Crude Est.	Kriens Peterse	Lavenberg et al.	Jackknife	Roberts (1978)
1	0.5	247,170	1,820	1,863	1,830	15,933
1	1.0	248,220	1,798	1,839	2,142	13,764
1	5.0	249,854	3,160	3,240	3,843	12,587
1	10.0	252,109	5,365	5,496	6,282	26,925
1	30.0	254,837	8,434	8,631	10,020	69,976
1M	1.0	157,253	1,741	1,779	1,844	13,604
1M	10.0	155,802	4,952	5,062	5,474	24,943
2	0.5	1,677,792	14,813	15,227	15,229	142,200
2	1.0	1,720,641	29,969	30,798	31,170	219,296
2	5.0	1,700,385	64,956	66,749	69,549	250,219
2	10.0	1,714,079	97,261	99,958	100,663	473,182
2	70.0	1,723,094	279,665	287,410	297,538	3,381,306
3	0.5	12,881,632	22,866	23,466	23,149	204,671
3	1.0	12,896,835	26,616	27,293	26,962	187,005
3	5.0	12,794,730	47,908	49,120	48,939	185,280
3	10.0	12,915,863	71,144	72,985	72,585	348,140
3	30.0	12,934,154	135,938	139,522	138,962	1,129,575
4	0.5	4,191,663	293,850	303,970	327,210	2,809,476
4	1.0	4,168,212	170,819	176,749	178,540	1,194,441
4	5.0	4,143,232	229,127	237,124	244,343	896,081
4	10.0	4,135,537	561,415	580,908	608,948	2,813,090
4	30.0	3,886,359	1,414,353	1,463,387	1,573,497	12,459,839

Comparing tables 3 and 4 shows the effects of increasing the sample size (n) from 100 to 200. The crude estimator $\bar{x}$ gives higher estimated coverages; the lengths of its confidence intervals clearly decrease (table 5 versus table 6); its intervals remain much longer than the other estimators' intervals (table 6). Roberts (1978) remains too conservative. Kriens and Peterse (1986) - and hence Lavenberg et al. (1982) - and jackknifing result in coverages that do not clearly benefit from sample size increase; these methods do clearly result in shorter confidence intervals as n increases; compare tables 5 and 6.

We saw (below eq. 3.7) that Kriens and Peterse (1986) conjecture that the normal approximation holds, if the number of non-zero differences is at least 20 ($m \geq 20$) (and if some more conditions hold). Therefore we also stratify the Monte Carlo results on m; see Tables 7 and 8. Note that the last column aggregates the preceding two columns. Moreover we aggregate the five error percentages per population; for population 1M we aggregate over two error percentages (so the maximum number of observations per stratum is 5×600). Let K_s denote the number of Monte Carlo observations in stratum s ($s = 1,2,3,4$). An asterisk (*) means that the estimated coverage in the stratum is not significantly different from the nominal $1-\alpha = 0.95$; this difference is tested at $\alpha' = 0.05$; unfortunately, a small number of observations K_s tends to create an asterisk (small power of test). We do not display the results for Lavenberg et al. (1982) because these results are very similar to Kriens and Peterse (1986).

Table 7 suggests that condition (ii) of Kriens and Peterse (1986) is too strong; their method also gives good results for $15 \leq m \leq 20$.

Table 7

Estimated Coverages after Stratification on Number of Non-zero Differences $\underline{m}$, for Sample Size n = 100

POPULATION 1 (number of obs.)	$5 < m \leq 10$ (585)	$10 < m \leq 15$ (243)	$15 < m \leq 20$ (42)	$10 < m \leq 20$ (285)
Crude $\bar{x}$	0.7624	0.8148	0.8095	0.8140
Kriens & Peterse	0.9111	0.9136	0.9524*	0.9193
Jackknifing	0.9521*	0.9506*	0.9762*	0.9544*
Roberts	1	1	1 *	1
POPULATION 1M	$5 < m \leq 10$ (333)	$10 < m \leq 15$ (212)	$15 < m \leq 20$ (24)	$10 < m \leq 20$ (236)
Crude $\bar{x}$	0.8438	0.9198	0.8750*	0.9153
Kriens & Peterse	0.9009	0.9104	0.9167*	0.9110
Jackknifing	0.9309*	0.9528*	0.9583*	0.9534*
Roberts	1	1	1 *	1
POPULATION 2	$5 < m \leq 10$ (561)	$10 < m \leq 15$ (240)	$15 < m \leq 20$ (31)	$10 < m \leq 20$ (271)
Crude $\bar{x}$	0.9144	0.8833	0.9355*	0.8843
Kriens & Peterse	0.9590*	0.9917	0.9677*	0.9890
Jackknifing	0.9697	1	0.9677*	0.9960
Roberts	0.9982	1	1 *	1
POPULATION 3	$5 < m \leq 10$ (559)	$10 < m \leq 15$ (244)	$15 < m \leq 20$ (34)	$10 < m \leq 20$ (278)
Crude $\bar{x}$	0.8318	0.8279	0.9706*	0.8454
Kriens & Peterse	0.7746	0.9344*	0.9118*	0.9316*
Jackknifing	0.7818	0.9426*	0.9118*	0.9496*
Roberts	0.9678*	0.9918	1 *	0.9928
POPULATION 4	$5 < m \leq 10$ (572)	$10 < m \leq 15$ (235)	$15 < m \leq 20$ (22)	$10 < m \leq 20$ (257)
Crude $\bar{x}$	0.9283	0.9362*	0.7727	0.9222
Kriens & Peterse	0.6329	0.7617	0.8182	0.7665
Jackknifing	0.6469	0.7745	0.8636*	0.7821
Roberts	0.8549	0.9489*	0.9091*	0.9455*

Table 8

Estimated Coverages after Stratification on Number of Non-zero Differences $\underline{m}$, for Sample Size n = 200

POPULATION 1 (number of obs.)	$5 < m \leq 10$ (333)	$10 < m \leq 15$ (308)	$15 < m \leq 20$ (252)	$10 < m \leq 20$ (560)
Crude $\bar{x}$	0.8198	0.8377	0.8333	0.8357
Kriens & Peterse	0.9459*	0.9156	0.9206	0.9174
Jackknifing	0.9760	0.9675*	0.9286*	0.9500*
Roberts	1	1	1	1
POPULATION 1M	$5 < m \leq 10$ (27)	$10 < m \leq 15$ (78)	$15 < m \leq 20$ (254)	$10 < m \leq 20$ (332)
Crude $\bar{x}$	0.8889*	0.8462	0.9095	0.8946
Kriens & Peterse	0.9630*	0.8333	0.9016	0.8855
Jackknifing	0.9630*	0.8718	0.9330*	0.8946
Roberts	1 *	1	1	1
POPULATION 2	$5 < m \leq 10$ (343)	$10 < m \leq 15$ (314)	$15 < m \leq 20$ (281)	$10 < m \leq 20$ (595)
Crude $\bar{x}$	0.9475*	0.9395*	0.9431*	0.9412*
Kriens & Peterse	0.8921	0.9650*	0.9573*	0.9614*
Jackknifing	0.9038	0.9745	0.9715*	0.9731
Roberts	0.9942	1	1	1
POPULATION 3	$5 < m \leq 10$ (319)	$10 < m \leq 15$ (317)	$15 < m \leq 20$ (281)	$10 < m \leq 20$ (598)
Crude $\bar{x}$	0.8652	0.8738	0.8790	0.8762
Kriens & Peterse	0.7712	0.8580	0.8968	0.8762
Jackknifing	0.7743	0.8644	0.9004	0.8798
Roberts	0.9687*	0.9969	1	0.9984
POPULATION 4	$5 < m \leq 10$ (322)	$10 < m \leq 15$ (304)	$15 < m \leq 20$ (286)	$10 < m \leq 20$ (590)
Crude $\bar{x}$	0.9348*	0.9016	0.9467*	0.9322
Kriens & Peterse	0.5807	0.7237	0,6748	0.7000
Jackknifing	0.5807	0.7237	0.6783	0.7017
Roberts	0.8758	0.9375*	0.9441*	0.9409*

Table 8 compared to Table 7, shows that as the sample size (n) increases, the number of non-zero differences ($\underline{m}$) increases too, which is obvious.

11. Conclusions

Monte Carlo studies are often based on arbitrary experimental designs. In audit sampling, however, Neter and Loebbecke (1975)'s populations are accepted as representative.

The question we addressed, is "may the auditor accept the total book value ($\sum_{1}^{N} y_i$) as a correct figure for the total audit value ($\sum_{1}^{N} x_i$)?" To save time the auditor starts with a sample (of size n) and observes both the book values $\underline{y}_j$ and the corresponding audit values $\underline{x}_j$ (j = 1,...,n). Some preliminary conditions on these n pairs $(\underline{x}_j, \underline{y}_j)$ are:

(i) If "nearly" all differences $\underline{x}_j - \underline{y}_j$ have the same sign, then the bookkeeping systematically overestimates or underestimates the true values, and the auditor will not simply accept the book values as a substitute for the audit values; he will follow up with additional audit procedures to find out whether there are acceptable reasons for this phenomenon. Kriens and Peterse (1986, p. 30) use the sign statistic to test if $P(\underline{x}_j > \underline{y}_j) = P(\underline{x}_j \leq \underline{y}_j) = 0.5$; we recommend such a test.

(ii) If all book values in the sample are equal to the audit values ($\forall i$: $\underline{x}_i = \underline{y}_i$), then the auditor needs no mathematical statistics to conclude that the books are extremely accurate and to accept the book values y_i.

If these two preliminary conditions are not met, then a $1-\alpha$ confidence interval for the audit value is needed.

The literature gives several point estimators and confidence intervals. Roberts (1970) gave a complicated procedure which, however, performs very poorly, according to our Monte Carlo results. Frost and Tamura (1982) investigated the ratio estimator, possibly combined with jackknifing, but they found several extremely low coverages such as $1-\hat{\alpha} = 0.30$. Beck (1980) examined regression estimators without jackknifing; our study indicates that jackknifing the regression estimators yields a more robust estimator (also see below).

The crude estimator $\bar{\underline{x}}$ does not yield better coverage than the various regression estimators do, except for populaton 4; then, however, the crude estimator gives confidence intervals so long that they are of no practical value. The regression estimators drastically reduce confidence interval lengths. The rule of thumb in Kriens and Peterse (1986), namely $m \geq 20$, can be relaxed, namely $m \geq 15$.

Practitioners want simple guidelines.We give the following rules of thumb:

(i) The use of auxiliary information, namely the book values y_i ($i = 1,\ldots,N$) and their sampled values $\underline{y}_j$ ($j = 1,\ldots,n$), yields a much

shorter 1-α confidence interval than the crude estimator $\bar{x}$ does. (The t statistic modified to account for skewness gives better coverage than the crude estimator does, but it presumably still gives long intervals compared to estimators based on auxiliary information.)

(ii) Estimators using auxiliary information, are the ratio, the difference, and the regression estimators. Their relative performances differ. Good relative performance is necessary, especially if the bookkeeping shows systematic error: the intercept δ is not equal to zero or the slope γ is not equal to one. Then regression estimators have the best relative performance.

(iii) Jackknifing improves the coverage of the regression estimator; yet that coverage is not acceptable if the error percentages are too low or the error distributions are too skew. Fortunately the sample $(\underline{x},\underline{y})$ gives a clear warning. If the error distribution is skew then most errors have the same sign and the auditor will proceed to additional audit procedures not based on sampling. If there are only a few non-zero differences, then the auditor may accept the book value. In all other situations the auditor uses the jackknifed regression estimator.

Appendix 1: List of major symbols

α : probability of type I error

β : probability of type II error

γ : slope of regression estimator

δ : intercept of regression estimator

μ_x : mean audit value ($= \vartheta/N = \vartheta'$)

μ_y : mean book value $(\sum_1^N y_i/N)$

ρ : correlation coefficient

σ : standard deviation

ϑ : total audit value $(\sum_1^N x_i)$

ϑ' : mean audit value $(\sum_1^N x_i/N)$

i : index in population ($i = 1,\ldots,N$)

j : index in sample ($j = 1,\ldots,n$)

$\underline{J}$: pseudo-value in jackknifing

$\underline{L}_{\alpha}$: lower bound of 1-α confidence interval

$\underline{m}$: number of non-zero differences in sample

n : sample size

N : population size

$t_v^{(\alpha)}$: 1-α quantile of Student statistic $\underline{t}$ with v degrees of freedom

$\underline{x}$: audit (true) value

$\bar{\underline{x}}$: crude estimator $(\sum_1^n \underline{x}_j/n)$

$\bar{\underline{x}}(\hat{\gamma})$: regression estimator

$\bar{\underline{x}}(\gamma)$: difference estimator

$\bar{x}_r$: ratio estimator

y : book value

$z^{(\alpha)}$: 1-α quantile of distribution of standard normal z

U_α : upper bound of 1-α confidence interval

W : matrix of explanatory variables

v_{11} : element (1,1) of $(WW')^{-1}$

v : binary variable to estimate coverage

Appendix 2: The Roberts (1978) procedure

Roberts (1978, pp. 90-91) uses the following symbols, listed in order of appearance:

Roberts' symbols	Our symbols
$\hat{X}_G$	$N\,\bar{x}(\hat{\gamma})$
b	$\hat{\gamma}$
Y	$N\,\mu_y = \sum_1^N y_i$
U_R	$z^{(\alpha)}$
R	1-α

He proposes the following procedure:

"Evaluating Results. The estimated audited amount using regression estimation is

$$\hat{X}_G = N\bar{x} + b(Y-N\bar{y}),$$

where b is the estimated regression coefficient based on the entire sample.

The achieved precision of the regression estimate of the total audited amount is calculated as follows:

$$A'_G = \frac{NU_R S_G}{\sqrt{n}} \sqrt{1-n/N}$$

where S_G is the estimated standard deviation of the regression population.

... the estimated standard deviation (S_G) may underestimate the population standard deviation when only a few differences are observed. ... The following procedure may be used to obtain an estimate of the standard deviation of the regression population in cases where the auditor feels the proportion of differences in the sample may be smaller than the proportion of differences in the population.

1. From a binomial (or Poisson) table determine the achieved upper precision limit for the proportion of differences based upon the observed number of differences (m), the sample size (n), and a desired one-sided reliability (R_1). Label this number $P_u(m)$.

2. Calculate the mean difference for only the sample units containing non-zero differences, as follows:

$$\bar{d}_m = \frac{\Sigma\, d_j}{m}.$$

3. Calculate the estimated standard deviation for the non-zero population difference as follows:

$$S_D(m) = \sqrt{\frac{\Sigma\, d_j^2 - m\bar{d}_m^2}{m-1}}.$$

4. Calculate the quantity

$$\frac{1}{n-2}\,\frac{(\Sigma\, d_j y_j - m\bar{d}_m\bar{y}_n)^2}{\Sigma\, y_j^2 - n\bar{y}^2} = \frac{(n-1)}{(n-2)}\,(b-1)^2 S_Y^2,$$

where $\bar{y}_n$ is the average recorded amount for the entire sample. [In the denominator a misprint occurs: $\bar{y}$ should be $\bar{y}_n$.]

5. Determine the estimated standard deviation of the regression population as follows:

$$\hat{\sigma}_G = \sqrt{P_U(m)S_D^2(m) + P_U(m)(1-P_U(m))\bar{d}_m^2 - \frac{(n-1)}{(n-2)}\,(b-1)^2 S_Y^2}.$$

In addition, when computing the achieved precision (A'_G), use the reliability factor from the Student's t-table with (m-2) degrees of freedom instead of te U_R based on the normal tables."

Appendix 3: The Lavenberg et al. (1982) model

Lavenberg et al. (1982, pp. 183-185) consider the following situation. Suppose we simulate a queuing system with $Q \geq 1$ service stations, each server with its own mean service time μ_q $(q = 1,\ldots,Q)$. The simulation samples the actual service times from the Q service time distributions which results in the average service times $\underline{c}_q$. The simulation run has as output $\underline{y}$ (for example, $\underline{y}$ denotes average waiting time at the end of the simulation run). Then they proceed as follows.

"Consider a single run of a simulation and let μ be an unknown quantity to be estimated. Let Y be an unbiased estimator of μ, generated from the simulation run. A random variable C, also generated from the simulation run, is a control variable if its expectation is known and if it is correlated with Y. Let C be a column vector of Q such control variables with elements $C_q : q = 1,\ldots,Q$. Let μ_C be the expectation of C, and μ_q the expectation of $C_q : q = 1,\ldots,Q$. Then for any column vector of constant coefficients, a, the estimator $Y(a) = Y - a'(C-\mu_C)$ is an unbiased estimator of μ. The vector α which minimizes the variance of Y(a) is given by...

$$\alpha' = \sigma_{YC}\Sigma_C^{-1}, \qquad (1)$$

where Σ_C is the covariance matrix of C and σ_{YC} is a row vector given by $\sigma_{YC} = (\text{Cov}[Y,C_1],\ldots,\text{Cov}[Y,C_Q])$. ...

We let K be the number of independent replications and we let Y_k be the value of the estimator Y, C_k be the value of C and C_{qk} be the value of C_q on the k-th replication. ...

We assume Y and $C_1,\ldots,C_Q$ to have a multivariate normal distribution. (As noted below, this assumption can be generalized somewhat.) The estimators and control variables we consider ... are sample means. Hence, the multivariate normal assumption is a reasonable one. ... Then the conditional distribution of Y given $C = c$, i.e., $C_q = c_q$, $q = 1,\ldots,Q$ is univariate normal with expectation

$$E[Y|C = c] = \mu + \alpha'(c-\mu_c),$$

where α is given by (1)...

Hence conditional on $C_k = c_k$, $k = 1,\ldots,K$, we have a classical regression problem with $Y = X'\beta + \epsilon$, $Y' = (Y_1,\ldots,Y_K)$, $\beta = \begin{bmatrix}\mu\\ \alpha\end{bmatrix}$ and

$$X' = \begin{bmatrix} 1 & c_{11} - \mu_1 & \cdots & c_{Q1} - \mu_Q \\ \vdots & \vdots & & \vdots \\ 1 & c_{1K} - \mu_1 & \cdots & c_{QK} - \mu_Q \end{bmatrix}$$

where ϵ is a vector of independent, normally distributed random variables with mean zero and common variance σ^2 ... Now let the least squared estimators of μ and α be $\hat{\mu}_1$ and $\hat{\alpha}_1$. Then

$$\text{Var}[\hat{\mu}_1] = s_{11}\sigma^2 \tag{4}$$

where s_{11} is the upper leftmost element of $(XX')^{-1}$. From regression theory an unbiased estimator of σ^2 is

$$\sigma^2 = (\Sigma_{k=1}^{K} Y_k^2 - \Sigma_{k=1}^{K}(\hat{\mu}_1+\hat{\alpha}_1(C_k-\mu_C))^2\}/(K-Q-1). \tag{5}$$

Finally, conditional on $C_k = c_k$, $k = 1,\ldots,K$, we can generate a confidence interval and we have

$$\text{Prob}(\hat{\mu}_1 - s_{11}^{\frac{1}{2}}\hat{\sigma}t_{K-Q-1}(1-\delta/2) \leq \mu$$

$$\leq \hat{\mu}_1 + s_{11}^{\frac{1}{2}}\hat{\sigma}t_{K-Q-1}(1-\delta/2)\} = 1-\delta \tag{6}$$

where $t_n(x)$ is the 100 x-th percentile of the t-distribution with n degrees of freedom. Since the right hand side of (6) is a constant this confidence statement holds unconditionally. Further since least squares estimates are unbiased, by a parallel argument, $\hat{\mu}_1$ is unconditionally unbiased. ...

The above confidence interval can be derived under the more general conditions that the conditional distribution of Y given C = c is univariate normal with an expectation $d + e'c = \mu + \alpha'(c-\mu_c)$ and $\Delta^2 = [\sigma^2]$ from which the above results follow."

The main problem when applying Lavenberg et al. (1982)'s results is the difference in notation. Therefore we give the following "translation":

Lavenberg et al.'s symbols	Our symbols
μ : quantity to be estimated	$\vartheta/N = \mu_x$
Y : unbiased estimator of μ (= $E(Y)$)	$\underline{x}$
C : vector of Q auxiliary variables	$\underline{y}$ ($Q = 1$)
μ_c : $E(C)$	μ_y
$Y(a) = Y - a'(C-\mu_c)$: regression estimator	$\underline{x}(\gamma) = \underline{x} - \gamma(\underline{y}-\mu_y)$
Σ_c : covariance matrix of C	σ_y^2
σ_{YC} : row vector with $\text{cov}(Y,C_q)$ ($q = 1,\ldots,Q$)	$\text{cov}(\underline{x},\underline{y})$
$\alpha = \sigma_{YC}\,\Sigma_C^{-1}$: optimal value of a	$\gamma = \dfrac{\text{cov}(\underline{x},\underline{y})}{\sigma_y^2}$
K : number of independent replications (observations)	n
$E(Y\|C=c) = \mu + \alpha'(c-\mu_c)$	$E(\underline{x}\|\underline{y}=y) = \mu_x + \gamma(\underline{y}-\mu_y)$
$Y = X'\beta + \varepsilon$ with $\beta' = (\mu,\alpha)$	$\underline{x} = W'\Gamma + \varepsilon = \Gamma' = (\delta,\gamma)$
$X' = \begin{bmatrix} 1 & c_{11}-\mu_1 & \cdots & c_{Q1}-\mu_Q \\ \vdots & & & \\ 1 & c_{1K}-\mu_1 & \cdots & c_{QK}-\mu_Q \end{bmatrix}$	$W' = \begin{bmatrix} 1 & y_1-\mu_y \\ \vdots & \\ 1 & y_n-\mu_y \end{bmatrix}$
$\varepsilon \sim NID(0,\sigma^2)$	
$\hat{\mu}_1$: OLS estimator of μ [the intercept]	$\hat{\underline{\delta}} = \bar{\underline{x}} - \hat{\underline{\gamma}}(\bar{\underline{y}}-\mu_y) = \bar{\underline{x}}(\hat{\underline{\gamma}})$
$\text{var}(\hat{\mu}_1) = s_{11}\sigma^2$ where	$\text{var}(\bar{\underline{x}}(\hat{\underline{\gamma}})) = v_{11}\sigma^2$ where
s_{11} : upper leftmost element of $(XX')^{-1}$	v_{11} : upper leftmost element of $(WW')^{-1}$

$\hat{\sigma}^2 = (\sum_1^K Y_k^2 - \sum_1^K (\hat{\mu}_1+\hat{\alpha}_1(C_k-\mu_c)^2\}/(K-Q-1)$	$\hat{\sigma}^2 = (\sum_1^n (\underline{x}_j^2 - \sum_1^n (\hat{\delta}+\hat{\gamma}(y_j-\mu_y)^2\}/(n-2)$
$P(\hat{\mu}_1 - s_{11}^{\frac{1}{2}}\ \hat{\sigma}\ t_{K-Q-1}^{(\alpha/2)} \le \mu \le$	$P(\bar{x}(\hat{\underline{\gamma}}) - \sqrt{v_{11}}\ \hat{\sigma}\ t_{n-2}^{(\alpha/2)}$
$..+ \quad .. \quad) = 1-\alpha$	$\le \vartheta' \le \ldots) = 1-\alpha$
[we replace δ by α]	

Appendix 4: Jackknifed regression estimators

We eliminate the j^{th} pair $(\underline{x}_j, \underline{y}_j)$ where $j = 1,\ldots,n$. Analogous to (3.4) we compute

$$\bar{x}(\hat{\underline{\gamma}}_{-j}) = \bar{\underline{x}}_{-j} + \hat{\underline{\gamma}}_{-j}(\mu_y - \bar{\underline{y}}_{-j}) \qquad (j = 1,\ldots,n) \qquad (A4.1)$$

where

$$\bar{\underline{x}}_{-j} = \sum_{j' \neq j} \underline{x}_{j'}/(n-1) \qquad (A4.2)$$

$$\bar{\underline{y}}_{-j} = \sum_{j' \neq j} \underline{y}_{j'}/(n-1) \qquad (A4.3)$$

and where $\hat{\underline{\gamma}}_{-j}$ is computed from the n-1 pairs $(\underline{x}_{j'}, \underline{y}_{j'})$. Obviously we have the point estimator $\bar{\underline{J}}$:

$$\bar{\underline{J}} = n\ \bar{x}(\hat{\underline{\gamma}}) - (n-1) \sum_1^n \bar{x}(\hat{\underline{\gamma}}_{-j})/n \qquad (A4.4)$$

Notes

1) The auditor may also increase the sample size n so that $\hat{var}(\bar{\underline{x}}) = \frac{s^2_{\underline{x}}}{n}(1 - \frac{n}{N})$ has a smaller expected value (and a smaller variance); moreover $t_{n-1}^{(\alpha/2)}$ decreases. The auditor may also increase α so that $t_{n-1}^{(\alpha/2)}$ decreases, but then $1-\alpha$ decreases or the type I error increases (while the type II error decreases). The main text shows more refined estimators which decrease the type II error while the type I error remains unchanged.

2) These data are available on magnetic tape from dr. John Neter, Professor of Management Sciences and Statistics, Department of Management Sciences and Information Technology, College of Business Administration, The University of Georgia, Athens, Georgia 30602 (phone 404-542-3585).

3) Obviously α' does not need to be identical to α. To test $1-\hat{\underline{\alpha}}$ we use the variable $\underline{v}$ of (10.1) which is binomial, conditional on the number of observations in stratum s (say) $\underline{K}_s$; that is, if H_0' : $E(\hat{\underline{\alpha}}) = \alpha$ then $var(1-\hat{\underline{\alpha}}) = \alpha(1-\alpha)/K_s$. The binomial distribution's critical value is approximated by the normal quantile $z^{(\alpha'/2)}$ where α' is the type I error rate when testing H_0' (we used the symbol H_0 in eq. 1.2 and tested H_0 with type 1 rate α).

References

BAKER, R.L. and R.M. COPELAND, Evaluation of the stratified regression estimator for auditing accounting populations. JOURNAL OF ACCOUNTING RESEARCH, 17, 1979, pp. 606-617.

BECK, P.J., A critical analysis of the regression estimator in audit sampling. JOURNAL OF ACCOUNTING RESEARCH, 18, no. 1, Spring 1980, pp. 16-37.

DENG, L.Y. AND C.F.J. WU, Estimation of variance of the regression estimator. JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 82, no. 398, June 1987, pp. 568-576.

FROST, P.A. and H. TAMURA, Jackknifed ratio estimator in statistical auditing. JOURNAL OF ACCOUNTING RESEARCH, 20, no. 1, Spring 1982, pp. 103-120.

KEEPING, E.S., INTRODUCTION TO STATISTICAL INFERENCE. D. Van Nostrand Company, Inc., Princeton, 1962.

KLEIJNEN, J.P.C., STATISTICAL TOOLS FOR SIMULATION PRACTITIONERS. Marcel Dekker, Inc., New York, 1987.

KLEIJNEN, J.P.C., P.C.A. KARREMANS, W.K. OORTWIJN and W.J.H. VAN GROENENDAAL, Jackknifing estimated weighted least squares: JEWLS. COMMUNICATIONS IN STATISTICS, THEORY AND METHODS, 16, no. 3, 1987, pp. 747-764.

KLEIJNEN, J.P.C., G.L.J. KLOPPENBURG and F.L. MEEUWSEN, Testing the mean of an asymmetric population: Johnson's modified t test revisited. COMMUNICATIONS IN STATISTICS, SIMULATION AND COMPUTATION, 15, no. 3, 1986, pp. 715-732.

KRIENS, J. and J.J.M. PETERSE, Toepassing van de regressieschatter in de accountantscontrole. (Application of the regression estimator in auditing.), KWANTITATIEVE METHODEN, 20, 1986, pp. 21-36.

LAVENBERG, S.S., T.L. MOELLER and P.D. WELCH, Statistical results on control variables with application to queueing network simulation. OPERATIONS RESEARCH, 30,no. 1, Jan.-Feb. 1982, pp. 182-202.

MILLER, R.G., The jackknife - a review. BIOMETRIKA, 61, 1974, pp. 1-15.

NETER, J. and J.K. LOEBBECKE, BEHAVIOR OF MAJOR STATISTICAL ESTIMATORS IN SAMPLING ACCOUNTING POPULATIONS: AN EMPIRICAL STUDY. American Institute of Certified Public Accountants, Inc., New York, 1975.

PATTERSON, R.L., Difficulties involved in the estimation of a population mean using transformed sample data. TECHNOMETRICS, 8, no. 3, August 1966, pp. 535-537.

ROBERTS, D.M., STATISTICAL AUDITING. American Institute of Certified Public Accountants, New York, 1978.

WILKS, S.S., MATHEMATICAL STATISTICS, John Wiley and Sons, Inc., New York, 1962.

IN 1987 REEDS VERSCHENEN

242 Gerard van den Berg
Nonstationarity in job search theory

243 Annie Cuyt, Brigitte Verdonk
Block-tridiagonal linear systems and branched continued fractions

244 J.C. de Vos, W. Vervaat
Local Times of Bernoulli Walk

245 Arie Kapteyn, Peter Kooreman, Rob Willemse
Some methodological issues in the implementation of subjective poverty definitions

246 J.P.C. Kleijnen, J. Kriens, M.C.H.M. Lafleur, J.H.F. Pardoel
Sampling for Quality Inspection and Correction: AOQL Performance Criteria

247 D.B.J. Schouten
Algemene theorie van de internationale conjuncturele en strukturele afhankelijkheden

248 F.C. Bussemaker, W.H. Haemers, J.J. Seidel, E. Spence
On (v,k,λ) graphs and designs with trivial automorphism group

249 Peter M. Kort
The Influence of a Stochastic Environment on the Firm's Optimal Dynamic Investment Policy

250 R.H.J.M. Gradus
Preliminary version
The reaction of the firm on governmental policy: a game-theoretical approach

251 J.G. de Gooijer, R.M.J. Heuts
Higher order moments of bilinear time series processes with symmetrically distributed errors

252 P.H. Stevers, P.A.M. Versteijne
Evaluatie van marketing-activiteiten

253 H.P.A. Mulders, A.J. van Reeken
DATAAL - een hulpmiddel voor onderhoud van gegevensverzamelingen

254 P. Kooreman, A. Kapteyn
On the identifiability of household production functions with joint products: A comment

255 B. van Riel
Was er een profit-squeeze in de Nederlandse industrie?

256 R.P. Gilles
Economies with coalitional structures and core-like equilibrium concepts

257 P.H.M. Ruys, G. van der Laan
Computation of an industrial equilibrium

258 W.H. Haemers, A.E. Brouwer
Association schemes

259 G.J.M. van den Boom
Some modifications and applications of Rubinstein's perfect equilibrium model of bargaining

260 A.W.A. Boot, A.V. Thakor, G.F. Udell
Competition, Risk Neutrality and Loan Commitments

261 A.W.A. Boot, A.V. Thakor, G.F. Udell
Collateral and Borrower Risk

262 A. Kapteyn, I. Woittiez
Preference Interdependence and Habit Formation in Family Labor Supply

263 B. Bettonvil
A formal description of discrete event dynamic systems including perturbation analysis

264 Sylvester C.W. Eijffinger
A monthly model for the monetary policy in the Netherlands

265 F. van der Ploeg, A.J. de Zeeuw
Conflict over arms accumulation in market and command economies

266 F. van der Ploeg, A.J. de Zeeuw
Perfect equilibrium in a model of competitive arms accumulation

267 Aart de Zeeuw
Inflation and reputation: comment

268 A.J. de Zeeuw, F. van der Ploeg
Difference games and policy evaluation: a conceptual framework

269 Frederick van der Ploeg
Rationing in open economy and dynamic macroeconomics: a survey

270 G. van der Laan and A.J.J. Talman
Computing economic equilibria by variable dimension algorithms: state of the art

271 C.A.J.M. Dirven and A.J.J. Talman
A simplicial algorithm for finding equilibria in economies with linear production technologies

272 Th.E. Nijman and F.C. Palm
Consistent estimation of regression models with incompletely observed exogenous variables

273 Th.E. Nijman and F.C. Palm
Predictive accuracy gain from disaggregate sampling in arima - models

274 Raymond H.J.M. Gradus
The net present value of governmental policy: a possible way to find the Stackelberg solutions

275 Jack P.C. Kleijnen
A DSS for production planning: a case study including simulation and optimization

276 A.M.H. Gerards
A short proof of Tutte's characterization of totally unimodular matrices

277 Th. van de Klundert and F. van der Ploeg
Wage rigidity and capital mobility in an optimizing model of a small open economy

278 Peter M. Kort
The net present value in dynamic models of the firm

279 Th. van de Klundert
A Macroeconomic Two-Country Model with Price-Discriminating Monopolists

280 Arnoud Boot and Anjan V. Thakor
Dynamic equilibrium in a competitive credit market: intertemporal contracting as insurance against rationing

281 Arnoud Boot and Anjan V. Thakor
Appendix: "Dynamic equilibrium in a competitive credit market: intertemporal contracting as insurance against rationing

282 Arnoud Boot, Anjan V. Thakor and Gregory F. Udell
Credible commitments, contract enforcement problems and banks: intermediation as credibility assurance

283 Eduard Ponds
Wage bargaining and business cycles a Goodwin-Nash model

284 Prof.Dr. hab. Stefan Mynarski
The mechanism of restoring equilibrium and stability in polish market

285 P. Meulendijks
An exercise in welfare economics (II)

286 S. Jørgensen, P.M. Kort, G.J.C.Th. van Schijndel
Optimal investment, financing and dividends: a Stackelberg differential game

287 E. Nijssen, W. Reijnders
Privatisering en commercialisering; een oriëntatie ten aanzien van verzelfstandiging

288 C.B. Mulder
Inefficiency of automatically linking unemployment benefits to private sector wage rates

289 M.H.C. Paardekooper
A Quadratically convergent parallel Jacobi process for almost diagonal matrices with distinct eigenvalues

290 Pieter H.M. Ruys
Industries with private and public enterprises

291 J.J.A. Moors & J.C. van Houwelingen
Estimation of linear models with inequality restrictions

292 Arthur van Soest, Peter Kooreman
Vakantiebestemming en -bestedingen

293 Rob Alessie, Raymond Gradus, Bertrand Melenberg
The problem of not observing small expenditures in a consumer expenditure survey

294 F. Boekema, L. Oerlemans, A.J. Hendriks
Kansrijkheid en economische potentie: Top-down en bottom-up analyses

295 Rob Alessie, Bertrand Melenberg, Guglielmo Weber
Consumption, Leisure and Earnings-Related Liquidity Constraints: A Note

296 Arthur van Soest, Peter Kooreman
Estimation of the indirect translog demand system with binding non-negativity constraints

IN 1988 REEDS VERSCHENEN

297 Bert Bettonvil
Factor screening by sequential bifurcation

298 Robert P. Gilles
On perfect competition in an economy with a coalitional structure

299 Willem Selen, Ruud M. Heuts
Capacitated Lot-Size Production Planning in Process Industry

300 J. Kriens, J.Th. van Lieshout
Notes on the Markowitz portfolio selection method

301 Bert Bettonvil, Jack P.C. Kleijnen
Measurement scales and resolution IV designs: a note

302 Theo Nijman, Marno Verbeek
Estimation of time dependent parameters in lineair models
using cross sections, panels or both

303 Raymond H.J.M. Gradus
A differential game between government and firms: a non-cooperative approach

304 Leo W.G. Strijbosch, Ronald J.M.M. Does
Comparison of bias-reducing methods for estimating the parameter in dilution series

305 Drs. W.J. Reijnders, Drs. W.F. Verstappen
Strategische bespiegelingen betreffende het Nederlandse kwaliteits-concept